

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:55 ; Search time 91.75 Seconds
(without alignments)
14.079 Million cell updates/sec

Title: US-09-331-631a-8_COPY_80_119
Perfect score: 225
Sequence: 1 PEDPQRYEECCQECRCQERQPOCCQRCIKRFEDEQQ 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	221	98.2	588	1 VCLB_GOSHI	P09801 gossypium h
2	142	63.1	605	1 VCLA_GOSHI	P09799 gossypium h
3	86	38.2	1905	1 TAGB_DICDI	P54683 dictyostell
4	83	36.9	429	1 APA4_MACFA	P33621 macaca fasc
5	83	36.9	1403	1 PRO_DROME	P29617 drosophila
6	82	36.4	339	1 TFD2_HUMAN	P20226 homo sapien
7	82	36.4	401	1 APA4_PAPAN	P28758 papio anubi
8	78	34.7	648	1 KAPC_DICDI	P34099 dictyostell
9	77.5	34.4	2124	1 Y192_HUMAN	P03074 homo sapien
10	77	34.2	758	1 YN38_YEAST	P03825 mus musculu
11	76.5	34.0	467	1 INVO_MOUSE	P48997 mus musculu
12	76.5	34.0	788	1 YKFA_YEAST	P35732 saccharomyc
13	76	33.8	905	1 SNE5_YEAST	P18480 saccharomyc
14	75	33.3	1023	1 CLOC_DROME	O61735 drosophila
15	75	33.3	1154	1 WCI1_NEUCR	O01371 neurospora
16	74	32.9	550	1 CCF_DROME	P41046 drosophila
17	73	32.4	1505	1 SIMA_DROME	O24167 drosophila
18	72.5	32.2	262	1 GDAL_WHEAT	P04721 triticum ae
19	72.5	32.2	307	1 GDAN_WHEAT	P18573 triticum ae
20	72	32.0	538	1 CK11_YEAST	P23291 saccharomyc
21	71	31.6	286	1 GDA0_WHEAT	P20863 triticum ae
22	71	31.6	1365	1 SUZ2_DROME	P25172 drosophila
23	70.5	31.3	186	1 GDA8_WHEAT	P04728 triticum ae
24	70.5	31.3	255	1 LP61_EIMTE	P15714 eimeria ten
25	70	31.1	305	1 HOG1_HORVU	P17990 hordeum vul
26	70	31.1	313	1 GDA7_WHEAT	P04727 triticum ae
27	70	31.1	319	1 GDA5_WHEAT	P04725 triticum ae
28	70	31.1	3828	1 TRX_DROVI	O24742 drosophila
29	69	30.7	47	1 AGRP_LUECY	P56568 luffa cylin
30	69	30.7	296	1 GDA6_WHEAT	P04726 triticum ae
31	69	30.7	924	1 CIR8_DROME	P17970 drosophila
32	69	30.7	1090	1 NIT4_NEUCR	P28349 neurospora
33	68	30.2	398	1 PF21_ARATH	O04088 arabidopsis

34	67.5	30.0	2703	1 NOTC_DROME	P07207 drosophila
35	67	29.8	966	1 SNN6_YEAST	P14922 saccharomyc
36	67	29.8	1319	1 MNL_HUMAN	O10571 homo sapien
37	67	29.8	1898	1 TRHY_HUMAN	P13709 homo sapien
38	67	29.8	2038	1 FSH_DROME	P13709 drosophila
39	66	29.3	291	1 GDA2_WHEAT	P04722 triticum ae
40	66	29.3	297	1 GDA4_WHEAT	P04724 triticum ae
41	66	29.3	468	1 YOC1_CAEEL	O09260 caenorhabd
42	66	29.3	623	1 DSH_DROME	P51140 drosophila
43	66	29.3	1023	1 GLT_DROME	P33438 drosophila
44	66	29.3	1073	1 HR38_DROME	P49869 drosophila
45	66	29.3	1329	1 FTSK_ECOLI	P46889 escherichia

ALIGNMENTS

RESULT 1	ID	VCLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989 (Rel. 10, Created)				
DT	01-MAR-1989 (Rel. 10, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;				
OC	Malvales; Malvaceae; Gossypium.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and				
RT	germination. XVII. cDNA and amino acid sequences of the members of				
RT	the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN				
CC	BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVULICIN, CONGLICININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M16891; AAA33071.1; -.				
DR	PIR: A30838; FMCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -.				
DR	PFAM: PF00546; Seedstore_7s; 1.				
KW	Seed storage protein; Signal.				
FT	STGNAL	1	25		
FT	CHAIN	26	588		
SQ	SEQUENCE	588 AA: 69729 MW: 635699829AB8ADEB CRC64:			

Query Match 98.2%; Score 221; DB 1; Length 588;
Best local similarity 97.5%; Pred. No. 1.3e-15;
Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PEDPQRYEECCQECRCQERQPOCCQRCIKRFEDEQQ 40
|||||
Db 80 PEDPQRYEECCQECRCQERQPOCCQRCIKRFEDEQQ 119

RESULT 2
VCLA_GOSHI STANDARD; PRT; 605 AA.
ID VCLA_GOSHI

```
AC P09799; (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DR VITILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
CC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borito K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
   germination. XIX. Sequences and genomic organization of the alpha
   globulin (vitilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -I FUNCTION: SEED STORAGE PROTEIN.
CC -I SUBCELLULAR LOCATION: CYTOSOL; MEMBRANE-BOUND VACUOLAR PROTEIN
   BODIES.
CC -I SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
   CONVICTILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSRP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR PfAM: PF00546; Seedstore_7s; 1.
DR Seed storage protein; Signal.
FT SIGNAL
FT     1      23
FT CHAIN          VITILIN GC72-A.
SO SEQUENCE        605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match           63.1%; Score 142; DB 1; Length 605;
Best Local Similarity 53.8%; Pred. No. 1.le-07;
Matches    21; Conservative    12; Mismatches    6; Indels    0; Gaps    0;

OY      2  EDPPRRYECCOECHQAEERDOPQCORCLKRFEEQQQ 40
       |||||:::|:|:|:| | :|:| | ::|:||||
Db      78  EDPGRRYQDCRHCGQEERRLRHPHGEGCRGDEYEQQQ 116

RESULT      3
TAGB_DICDI TAGB_DICDI STANDARD; PRT; 1905 AA.
AC AC         PS4683;
RX MEDLINE=95262903; PubMed=7744252;
RA Shaubsky G., Kuspa A., Loomis W.F.;
RT "A multidrug resistance transporter/serine protease gene is required
   for prestalk specialization in Dictyostelium".
RL Genes Dev. 9:1111-1122(1995).
CC -I FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
   INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC -I SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO PEPTIDASE FAMILY
   S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
```

CC	-I- SIMILARITY: IN THE C-TERMINAL SECTION: BELONGS TO THE ATP-BINDING
CC	TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC	-I- SIMILARITY: STRONG, TO TAGC.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: U20432; AAA62212.1; -
DR	HSPD, P13569; INBD.
DR	DICTYDB: DD02059; TAGB.
DR	INTERPRO: IPR000209; -
DR	INTERPRO: IPR001140; -
DR	INTERPRO: IPR001617; -
DR	PFAM: PF00664; ABC_membrane.1.
DR	PFAM: PF00003; ABC_tran.1.
DR	PRINTS: PR00723; SUBTILISIN.
DR	PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
DR	PROSITE: PS00137; SUBTILASE_HIS.1.
DR	PROSITE: PS00138; SUBTILASE_SER.1.
DR	PROSITE: PS00211; ABC_TRANSPORTER.1.
DR	Hydrophobic; Serine protease; ATP-binding; Transport; Transmembrane;
KW	Signal.
FT	SIGNAL 1 ?
FT	CHAIN 1 1905 ?
FT	DOMAIN 378 700
FT	DOMAIN ? 1905
FT	TRANSMEM 1011 1031
FT	TRANSMEM 1076 1096
FT	TRANSMEM 1121 1141
FT	TRANSMEM 1210 1230
FT	TRANSMEM 1309 1329
FT	TRANSMEM 1332 1352
FT	ACT_SITE 387 387
FT	ACT_SITE 432 432
FT	ACT_SITE 695 695
FT	NP_BIND 1553 1560
FT	DOMAIN 63 67
FT	DOMAIN 95 104
FT	DOMAIN 107 134
FT	DOMAIN 311 321
FT	DOMAIN 833 839
FT	DOMAIN 838 844
FT	DOMAIN 871 876
FT	DOMAIN 1012 1015
FT	DOMAIN 1386 1389
FT	DOMAIN 1398 1404
FT	DOMAIN 1445 1450
FT	DOMAIN 1765 1779
FT	DOMAIN 1782 1785
FT	DOMAIN 1807 1812
FT	DOMAIN 1815 1860
FT	DOMAIN 1872 1878
FT	CARBOHYD 594 594
FT	CARBOHYD 621 621
FT	CARBOHYD 672 672
FT	CARBOHYD 747 747
FT	CARBOHYD 823 823
FT	CARBOHYD 1172 1172
FT	CARBOHYD 1522 1522
FT	CARBOHYD 1658 1658
SO	SEQUENCE 1905 AA; 212518 MW; BBE223FA8B9AE13C CRC64;

Query Match 38.2%; Score 86; DB 1; Length 1905;

Best Local Similarity 42.5%; Pred. NO. 0.12;

Matches 17; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

1 PEDPQRYEECCQECRCQDERQPCQCRCKLRFQEDQDQ 40


```

QY 2 EDPORRYECQOECCROOEHOOPQCCQRCRKRFEEQ00 40
DB 53 EEOORROO0000000000000000000000000000 91

RESULT 7
AP04_PAPAN STANDARD: PRT: 401 AA.
AC 028758;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DE 15-DEC-1998 (rel. 37, Last annotation update)
DE APOLOPORTEIN A-IV PRECURSOR (APO-ATV) (FRAGMENT).
GN APO04.
OS Papio anubis (Olive baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=93340170; PubMed=8101842;
RA Hixson J.E., Kammerer C.M., Mott G.E., Britten M.L., Birnbaum S.,
RA Powers P.K., Vandenberg J.L.; Identification of Lys-76-->Glu that
RA "Baboon apolipoprotein A-IV, identification of Lys-76-->Glu that
RT distinguishes two common isoforms and detection of length
RT polymorphisms at the carboxyl terminus."
RL J. Biol. Chem. 268:15667-15673(1993).
CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CARBOHYDRATE REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPIASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED PRIMARILY IN THE INTESTINE.
CC SECRETED IN PLASMA.
CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -1- POLYMORPHISM: THERE ARE TWO COMMON APOA-IV ISOFORMS, I (SHOWN
CC HERE) AND E. THE I ISOFORM IS ASSOCIATED WITH HIGHER LEVELS OF
CC HIGH DENSITY LIPOPROTEIN-C ON A HIGH CHOLESTEROL, SATURATED FAT
CC DIET.
CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13174; AAA35379.1; -.
DR HSSP: P02649; INFO.
DR INTERPRO: IPR000074; -.
DR PFAM: PF01442; Apolipoprotein: 1.
KM Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal;
KW POLYMORPHISM.
FT NON_TER 1 1
FT SIGNAL. <1 4
FT CHAIN 5 401 POTENTIAL.
FT DOMAIN 17 314 APOLOPORTEIN A-IV.
FT REPEAT 17 38 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 44 65 1.
FT REPEAT 66 87 2.
FT REPEAT 99 120 3.
FT REPEAT 121 142 4.
FT REPEAT 143 164 5.
FT REPEAT 165 186 6.
FT REPEAT 187 208 7.

```

```

FT REPEAT 209 230 9.
FT REPEAT 231 252 10.
FT REPEAT 253 270 11.
FT REPEAT 271 292 12.
FT REPEAT 293 314 13.
FT DOMAIN 356 394 GLU/GLN-RICH.
FT VARIANT 80 80 K->E (IN ISOFORM E).
SQ SEQUENCE 401 AA; 46538 MW; 0A76D1284AA9837F CRC64;

Query Match 36.4%; Score 82; DB 1; Length 401;
Best Local Similarity 47.5%; Pred. No. 0.079;
Matches 19; Conservative 6; Mismatches 9; Indels 6; Gaps 1;

QY 1 PEDPORRYECQOECCROOEHOOPQCCQRCRKRFEEQ00 40
DB 357 PQOQOEQOQOEQOQOEQOQOEQOQO-----PQEQ00 390

RESULT 8
KAPC_DICDI STANDARD: PRT: 648 AA.
AC P34099;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DE 30-MAY-2000 (rel. 39, Last annotation update)
DE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT (EC 2.7.1.37).
GN PKAC OR PK2 OR PK3.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=91323730; PubMed=1864510;
RA Buerki E., Anjard C., Scholder J.-C., Raymond C.D.;
RT "Isolation of two genes encoding putative protein kinases regulated
RT during Dictyostelium discoideum development."
RL Gene 102:57-65(1991).
RN [2]
RX CHARACTERIZATION.
RX MEDLINE=93385090; PubMed=8373760;
RA Anjard C., Etcheverne L., Pinard S., Veron M., Raymond C.D.;
RT "An unusual catalytic subunit for the CAMP-dependent protein kinase
RT of Dictyostelium discoideum."
RL Biochemistry 32:9532-9538(1993).
RN [3]
RP CHARACTERIZATION.
RC STRAIN=AX3;
RX MEDLINE=93066311; PubMed=1332055;
RA Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;
RT "DgkK3, which plays essential roles during Dictyostelium development,
RT encodes the catalytic subunit of CAMP-dependent protein kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC -1- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.
CC -1- SUBUNIT: IN DICTYOSTELIUM THE HOMOZYME IS A DIMER COMPOSED OF
CC A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC -1- DEVELOPMENTAL STAGE: CAMP ACTIVITY IS LOW IN VEGETATIVELY GROWING
CC AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES
CC A MAXIMUM AT CULMINATION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAMP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38703; -. NOT_ANNOTATED_CDS.
DR PIR: J01150; J01150.

```



```
FT DOMAIN 489 588 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 714 882 PRO-RICH.
FT DOMAIN 755 798 ARG/LYS-RICH (BASIC).
FT CONFLICT 564 564 E -> D (IN REF. 1).
SO SEQUENCE 905 AA: 102557 MW: A287B4A648DD1A35 CRC64;

Query Match 33.8%; Score 76; DB 1; Length 905;
Best Local Similarity 35.9%; Pred. No. 0.63;
Matches 14; Conservative 16; Mismatches 9; Indels 0; Gaps 0;

Oy 2 EDPORRECOEQRQOEORQPOCQCRKRFEEQOO 40
Db 225 00000000000000000000000000000000 263

RESULT 14
CLOC_DROME STANDARD; PRT: 1023 AA.
ID CLOC_DROME 061735; 076342; 077137;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CIRCADIAN LOCOMOTOR OUTPUT CYCLES KAPUT PROTEIN (CLOCK) (DPA1).
GN CLK OR JRK OR CLOCK OR PAS1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE-HEAD;
RX MEDLINE=98279147; PubMed=9616122;
RA Darlington T.K., Mager-Smith K., Ceriani M.F., Stekuns D., Gekakis N.,
RA Steeves T.D.L., Weitz C.J., Takahashi J.S., Kay S.A.:
RT "Closing the circadian loop: Clock-induced transcription of its own
RT inhibitors per and tim.";
RL Science 280:1599-1603(1998).
RN 121
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC TISSUE-HEAD;
RX MEDLINE=98292177; PubMed=9630223;
RA Allada R., White N.E., So W.V., Hall J.C., Rosbash M.;
RT "A mutant Drosophila homolog of mammalian Clock disrupts circadian
RT rhythms and transcription of period and timeless.";
RL Cell 93:791-804(1998).
RN 131
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=98414630; PubMed=9742131;
RA Bae K., Lee C., Sidote D., Chuang K.-Y., Egerly I.;
RT "Circadian regulation of a Drosophila homolog of the mammalian clock
RT gene: PER and TIM function as positive regulators.";
RL Mol. Cell Biol. 18:6142-6151(1998).
CC -1- FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR
CC AND GENERATES A RHYTHMIC OUTPUT WITH A PERIOD OF ABOUT 24 HOURS.
CC OSCILLATES IN ANTIPHASE TO THE CYCLING OBSERVED FOR PERIOD (PER)
CC AND TIMELESS (TIM). ACCORDING TO REF.3, REACHES PEAK ABUNDANCE
CC WITHIN SEVERAL HOURS OF THE DARK-LIGHT TRANSITION AT ZTO
CC (ZEITGEBER 0), WHEREAS REF.1 DESCRIBES BIMODAL OSCILLATING
CC EXPRESSION WITH MAXIMUM AT ZT5 AND ZT23. CLOCK-CYCLE HETERODIMERS
CC ACTIVATE CYCLING TRANSCRIPTION OF PER AND TIM BY BINDING TO THE E-
CC BOX (3'-CAGCTG-5') PRESENT IN THEIR PROMOTERS, ONCE INDUCED,
CC PERIOD AND TIMELESS BLOCK CLOCK'S ABILITY TO TRANSACTIVATE THEIR
CC PROMOTERS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BLH PROTEIN. FORMS A HETERODIMER WITH CYCLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; THE FULL-LENGTH VARIANT A
CC (SHOWN HERE) AND VARIANT B: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC VARIANT B ENCODES TWO CONCEPTUAL PROTEINS, THE FIRST CONSISTS ONLY
CC OF THE BLH DOMAIN. THE OTHER CONSISTS OF THE PAS-1 AND ALL C-
CC TERMINAL DOMAINS. VARIANT B IS EXPRESSED WEAKLY AT ALL THE TIME OF
```

```
CC THE DAY, AND IT CYCLES IN PHASE WITH THE FULL-LENGTH FORM.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEAD, BODY, AND
CC APPENDAGE FRACTIONS.
CC -1- DOMAIN: CONTAINS THREE POLYGLUTAMINE REPEATS WHICH COULD
CC CORRESPOND TO THE TRANSACTIVATION DOMAIN. THE LENGTH OF THE
CC REPEATS IS POLYMORPHIC. IN THE ARRHYTHMIC MUTANT JRK, DELETION OF
CC THIS REGION LEADS TO THE LOSS OF CIRCADIAN RHYTHMICITY AND ALTERED
CC LIGHT RESPONSE.
CC -1- POLYMORPHISM: THE VARIABILITY IN LENGTH OF THE POLYGLUTAMINE
CC STRETCH IS DUE TO POLYMORPHISM OF THIS REGION.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067207; AAD10630.1; -
DR EMBL: AF065133; AAC39101.1; -
DR EMBL: AF069997; AAC62234.1; -
DR FLYBASE: FBgn0023076; clk.
DR INTERPRO: IPR000014; -
DR INTERPRO: IPR001067; -
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFM: PF00989; PAS; 2.
DR PFM: PF00989; PAS; 2.
DR PRINTS: PR00785; NCTRNLOCATR.
DR PROSITE: PS00038; HELIX LOOP HELIX; 1.
KW Transcription regulation; Nuclear protein; Repeat; Biological rhythms;
KW DNA-binding; Alternative splicing.
FT DOMAIN 12 24 BASIC DOMAIN.
FT REPEAT 86 152 PAS-1.
FT REPEAT 251 317 PAS-2.
FT DOMAIN 548 559 POLY-GLN.
FT DOMAIN 766 769 POLY-GLN.
FT DOMAIN 874 876 POLY-GLN.
FT DOMAIN 877 895 POLY-ASN.
FT DOMAIN 953 963 POLY-ASN.
FT DOMAIN 976 1023 POLY-GLN.
FT VARIAT 816 823 IMPLICATED IN THE CIRCADIAN RHYTHMICITY.
FT CONFLICT 12 12 MISSING.
FT CONFLICT 32 32 K -> KSFLLC (IN REF. 3).
FT CONFLICT 128 128 N -> D (IN REF. 3).
FT CONFLICT 128 128 N -> K (IN REF. 2).
FT CONFLICT 555 555 N -> S (IN REF. 1).
FT CONFLICT 605 605 I -> L (IN REF. 3).
FT CONFLICT 912 912 Y -> C (IN REF. 3).
SO SEQUENCE 1023 AA: 115751 MW: 514374C8C050DAFB CRC64;

Query Match 33.3%; Score 75; DB 1; Length 1023;
Best Local Similarity 38.9%; Pred. No. 0.88;
Matches 14; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

Oy 5 ORRYECCOECRQOEORQPOCQCRKRFEEQOO 40
Db 794 00000000000000000000000000000000 829

RESULT 15
WCL_NEUCR STANDARD; PRT: 1154 AA.
ID WCL_NEUCR 001371;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-NOV-2000 (Rel. 39, Last annotation update)
```


GN WHITE COLLAR 1 PROTEIN (WC1).
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-74-OR23-1A;
 RX MEDLINE=96203083; PubMed=8612589;
 RA Ballario P., Vittorioso P., Magrelli A., Talora C., Cabibbo A.,
 RA Macino G.;
 RT "White collar-1, a central regulator of blue light responses in
 RT Neurospora, is a zinc finger protein.";
 RL EMO J. 15:1650-1657(1996).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR INVOLVED IN LIGHT
 CC REGULATION. BINDS AND AFFECTS BLUE LIGHT REGULATION OF THE AL-3
 CC GENE. WC1 AND WC2 PROTEINS INTERACT VIA HOMOLOGOUS PAS DOMAINS,
 CC BIND TO PROMOTERS OF LIGHT REGULATED GENES SUCH AS PRO, AND
 CC ACTIVATE TRANSCRIPTION.
 CC -1- SUBUNIT: HETERODIMER OF WC1 AND WC2 (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- INDUCTION: BY BLUE LIGHT.
 CC -1- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING
 CC GENE EXPRESSION.
 CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
 CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on ways
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
 CC or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).

 CC EMBL; X94300; CAA63964.1; -.
 DR HSSP; P17678; 1GNV.
 DR INTERPRO; IPR000014; -.
 DR INTERPRO; IPR000679; -.
 DR INTERPRO; IPR001610; -.
 DR PFAM; PF00320; GATA.1.
 DR PFAM; PF00785; PAC; 1.
 DR PFAM; PF00989; PAC; 3.
 DR PROSITE; PS00344; GATA_2N_FINGER_1; 1.
 DR PROSITE; PS50114; GATA_2N_FINGER_2; 1.
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KW Nuclear protein; Repeat.
 FT DOMAIN 16 61 GLN-RICH.
 FT REPEAT 381 448 PAS-1.
 FT DOMAIN 469 508 PAC MOTIF.
 FT REPEAT 576 642 PAS-2.
 FT DOMAIN 650 691 PAC MOTIF.
 FT REPEAT 695 761 PAS-3.
 FT ZN_FING 935 960 GATA-TYPE.
 FT DOMAIN 21 57 POLY-GLN.
 FT DOMAIN 329 333 POLY-PRO.
 SEQUENCE 1154 AA; 125944 MW; CBA0CB6047ECCDC5 CRC64;

	Query Match	33.3%	Score 75;	DB 1;	Length 1154;
	Best Local Similarity	38.5%;	Pred. No. 0.98;		
	Matches	15;	Mismatches	11;	Indels
				Gaps	0;
Oy	1	PEDDPQRYRECCOECRCRODEROOPOCCOORCLKRFEEEOO	39		
	:	: :	: :	: :	: :
db	12	PEELDHOMHNOHOOOOONOOOONOOOONOOOONOOOONOOO	50		

Search completed: March 1, 2001, 16:16:57
Job time: 437 sec

